REMARKS

Reconsideration of the present application in view of the above amendments and the following remarks is requested respectfully.

Discussion of the Claims

Claims 1-56 are pending. Claims 12-19 and 26-56 have been previously withdrawn. Claims 1, 2, 3, 4, 6, 7, 11, 22 and 24 have been amended. Claims 57 and 58 has been added. Accordingly, Claims 1-11, 20-25, and 57 are presented for examination.

Summary of the Examiner's Action

Claims 1 and 21-25 stand rejected under 35 U.S.C. § 101 because the claimed invention is directed to non-statutory subject matter.

Claims 1, 3, 6, 8, 10, and 20-25 stand rejected under 35 U.S.C. § 112, first paragraph as failing to comply with the written description requirement.

Claims 1, 3, 6, 8, 10, 20-23 stand rejected under 35 U.S.C. § 102(b) as being anticipated by Sleep et al., 1990, Biotechnology, 8:42-46.

Examiner Interview

Applicants wish to thank Examiners Gupta and Gudibande for taking the time to discuss the present application and the pending rejections with Applicants' representative on October 17, 2007.

During the interview, the following were discussed: (1) election/restriction requirement, (2) statutory subject matter rejection, (3) the written description rejection and (4) the anticipation rejection.

With respect to the statutory subject matter and anticipation rejections, it was acknowledged that the cited motifs TILTA and TIASI do not fall within the scope of the motif recited in Claim 1, and thus these rejections will be withdrawn. Because of the misinterpretation of the scope of Claim 1, it was agreed that the next Office Action would not be made final.

Regarding the election/restriction requirement, it was confirmed that SEQ ID NO: 28 (the elected species) which is recited in Claim 11 is free of the prior art. Accordingly, the Examiners agreed to consider the allowability of Claim 11 if rewritten in independent form. As discussed above, because the scope of Claim 1 was misinterpreted, the Examiners will again search the motif as recited in Claim 1. If the claimed motif is free of the art, it was agreed that claims 2, 4, 5, 7 and 9 will be examined. Applicants agreed to add sequence identifiers to the claims that recite specific sequences.

With respect to the written description rejection, the Examiners agreed to reconsider the rejection if Applicants could show that the terms leader sequence and pre sequence are well known in the art.

Election/Restriction

The Examiner has indicated that elected species SEQ ID NO: 28 has been found free of art. The Examiner has withdrawn Claims 2, 4, 5, 7, 9, and 11 as being drawn to a non-elected species. Applicants respectfully disagree. The elected species SEQ ID NO: 28 is recited in Claim 11 and includes the pentapeptide motif Phe-Ile-Val-Ser-Ile (FIVSI), the motif specifically recited in Claim 7. As discussed above, Claim 11 was agreed to be drawn to the elected species. Moreover, if the Examiner finds that the motif as recited in Claim 1 is free of the art, Claims 2, 4, 5, 7 and 9 will be examined.

Discussion of the Applicant's Invention and Amendments

Claim 1 as amended and its dependent claims are directed to a polypeptide comprising a leader sequence which comprises a pre sequence and the X_1-X_2-X_3-X_4-X_5 motif as defined in the claim, and a mature protein. Applicant has unexpectedly found that the yield of secreted protein can be increased by providing the recited amino acid sequence motif in the leader sequence. Claim 1 was amended to clarify that the desired protein is a mature protein which is defined as a

protein without its pre sequence or pre-pro sequence. Support for this claim amendment can be found in the specification, for example, at page 18, lines 21-25.

As suggested by the Examiners during the Examiner interview, Claims 1, 2, 3, 4, 6, 7, and 11 have been amended to add sequence identifiers.

Claims 22 and 24 have been amended to correct dependencies and to be consistent with the amendments of Claim 1.

Claim 57 has been added to claim another embodiment of Applicant's invention. Support for this new claim can be found in the specification, for example, at page 3, lines 11-19. Claim 58 is directed to the particular elected species of that embodiment. Support for this claim can be found in the specification, for example, at page 22, lines 19-28 and original Claim 11.

Non-Statutory Subject Matter Rejection

On page 3 of the Action, the Examiner has rejected Claims 1 and 21-25 under 35 U.S.C. § 101 as being directed to non-statutory subject matter. Applicants submit respectfully that such rejection is misplaced.

35 U.S.C. § 101 defines patentable subject matter as "any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof." Here, Claim 1 and its dependent claims are directed to statutory subject matter – a manufacture or a composition of matter, and it appears that the Examiner is making more of an anticipation argument.

The Examiner has argued that the protein sequence of serum albumin (GenBank AAX82486) exhibits the motif TILTA at position 4-8 of the sequence and therefore Applicant's claimed polypeptide is a product of nature. TILTA represents threonine-isoleucine-leucine-threonine-alanine. The sequence of Claim 1, X_1 X_2 X_3 X_4 X_5, does not provide for a threonine at position X_1 (phenylalanine, tryptophan, or tyrosine), and thus the sequence TILTA differs from the presently claimed sequence due to the presence of this first threonine residue. Accordingly, this § 101 rejection should be withdrawn.

Written Description Rejection

On page 4 of the Action, the Examiner has rejected claims 1, 3, 6, 8, 10, and 20-25 under 35 U.S.C. § 112, first paragraph, as failing to comply with the written description requirement. The Examiner asserts that "the claim does not recite the nature of the 'leader sequence', 'pre sequence' or 'the protein heterologous to the leader sequence' in terms of the amino acid sequences that would properly define each of these different peptides" and "that the claims do not adequately provide structural characteristics" for these elements.

The written description requirement does not require a description of the complete structure of every species within a chemical genus. See *Utter v. Hiraga*, 845 F.2d 993, 998 (Fed. Cir. 1988). In *Enzo Biochem, Inc. v. Gen-Probe Inc.*, 296 F.3d 1316, 1324 (Fed. Cir. 2002), the Federal Circuit made clear that the written description requirement can be satisfied in a number of ways by disclosing, for example, "complete or partial structure, other physical and/or chemical

properties, functional characteristics when coupled with a known or disclosed correlation between function and structure, or some combination of characteristics."

As recognized by the Examiner, "the level of skill and knowledge in the art" is relevant to the determination of compliance with written description requirement. Here, the terms "leader sequence," "pre sequence," and, as amended, "a mature desired protein" are easily understandable to a person of ordinary skill in the art.

As amended, Claim 1 reads as follows:

1. A polypeptide comprising

(i) a leader sequence, the leader sequence comprising

(a) a secretion pre sequence, and

(b) the following motif :

$-X_1-X_2-X_3-X_4-X_5-$

where X_1 is phenylalanine, tryptophan, or tyrosine, X_2 is isoleucine, leucine, valine, alanine or methionine, X_3 is leucine, valine, alanine or methionine, X_4 is serine or threonine and X_5 is isoleucine, valine, alanine or methionine ; and

(ii) a mature desired protein.

Accordingly, Applicant claims a polypeptide comprising (i) a leader sequence and (ii) a mature protein. The leader sequence comprises (a) a secretion pre sequence and (b) the defined $-X_1-X_2-$

X_3-X_4-X_5- motif. Each of these elements is easily understood to one of skill in the art and discussed in the specification.

"Leader sequence" is defined at page 1, line 29, to page 2, line 1, as sequences which direct the secretion of proteins including pre sequences and pre-pro sequences. As can be seen from the attached pages from Voet and Voet, *Biochemistry*, 2nd Ed., 1995, John Wiley & Sons, pp. 1008-09, one of skill in the art recognizes that such "leader sequences" contain pre sequences or both pre and pro sequences. Similarly, at page 14, lines 9-12, "leader sequence" is defined functionally as a sequence that "causes more of that polypeptide to be secreted from the host cell in which it is produced." The function and structural properties of "pre sequences," also known as signal peptides, are well known in the art as evidenced, for example, by Gierasch, 1989, *Biochemistry*, 28(3), 923-931 (copy attached).

The -X_1-X_2-X_3-X_4-X_5- motif is defined structurally in Claim 1 as any combination of five amino acids where X_1 is phenylalanine, tryptophan, or tyrosine, X_2 is isoleucine, leucine, valine, alanine or methionine, X_3 is leucine, valine, alanine or methionine, X_4 is serine or threonine and X_5 is isoleucine, valine, alanine or methionine. The pre sequence may include all or part of the -X_1-X_2-X_3-X_4-X_5- motif. See specification at page 13, lines 9-18 and original Claim 20.

The term "mature desired protein" is defined as the secreted protein without its secretion pre sequence or the pre-pro sequence. See page 18, lines 21-25 and page 42, lines 8-11.

As explained above, Applicant has unexpectedly discovered that the claimed X_1-X_2-X_3-X_4-X_5- motif increases the yield of secreted protein. To the extent the Examiner is arguing that

Applicant does not have support for each of the pentapeptides within the claimed motif, Applicant submits that the claimed motif variants represent functional modifications to the exemplified motif (FIVSI), in that they allow for the presence of conservative amino acid substitutions at each of the positions of the exemplified motif:

- the first position of the motif, X_1, may only be an aromatic amino acid (like the exemplified Phe residue);
- X_2, X_3, and X_5 are selected from groups of amino acids which have non-polar side chains (like the exemplified Ile, Val and Ile residues, respectively);
- X_4 may only be either the exemplified Ser residue or its functional equivalent Thr, both of which have hydroxyl group-containing uncharged polar side chains.

It is well established that there is generally a low level of sequence identity between the amino acid sequences of different leader sequences and that their properties as leader sequences are determined by the chemical and steric properties of the component amino acids, rather than their absolute identities. See Gierasch, p. 27 (attached) ("Signal sequences seem likely to interact with many cellular components . . . but they apparently do so by virtue of their overall properties (residue type and patterns of residues) as opposed to specific sequence.").

Accordingly, the written description rejection should be withdrawn.

Anticipation Rejection

On page 7 of the Office Action, the Examiner rejected claims 1, 3, 6, 8, 10, and 20-23 under 35 U.S.C. §102(b) as being anticipated by Sleep et al., *Biotechnology*, 8, 42-46 (1990) (hereinafter "Sleep"). In presenting this rejection, the Examiner has relied on the presence of the sequence motif TIASI in sequence A of Sleep. TIASI represents the amino acid sequence threonine-isoleucine-alanine-serine-isoleucine. This does not fall within the claimed X_1-X_2-X_3-X_4-X_5- motif as defined in claim 1 as X_1 can not be threonine. Accordingly, Claim 1, and all claims depending therefrom, are patentable over Sleep.

Conclusion

In view of Applicant's claim amendment and the arguments presented above, the present application is believed to be in condition for allowance and an early notice thereof is earnestly solicited. Applicants request that the Examiner contact the undersigned before issuing another action.

Respectfully submitted,

/Marc S. Segal/
Marc S. Segal
Reg. No. 40,163

Synnestvedt & Lechner LLP
1101 Market Street, Suite 2600
Philadelphia, PA 19107
Telephone: (215) 923-4466
Facsimile: (215) 923-2189

Biochemistry

 Volume 28, Number 3, February 7, 1989

Perspectives in Biochemistry

Signal Sequences[†,‡]

Lila M. Gierasch

Departments of Pharmacology and Biochemistry, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75235-9041

Received October 7, 1988; Revised Manuscript Received November 7, 1988

While considerable progress has been made in the last 15 years in elucidating the mechanism of protein secretion [for reviews, see Verner and Schatz (1988), Randall et al. (1987), Briggs and Gierasch (1986), Rapoport (1986), Wickner and Lodish (1986), Walter and Lingappa (1986), and Walter et al. (1984)], the roles of the signal sequence are still poorly understood. Ironically, these 15–30 residue long, highly hydrophobic sequences constitute the most general requirement for export of a protein whether from yeast, higher eukaryotes, or bacteria. Several lines of evidence argue that signal sequences from these various organisms work in much the same way. Many features of the export pathway appear to be shared by all species, since most exported proteins can be translocated and processed correctly by the export machinery from several organisms [for an example, see Mueller et al. (1982); for an exception, see Bird et al. (1987)]. Recombinant proteins composed of a signal sequence from one organism and a mature secretory protein from another organism are frequently export competent (Yost et al., 1983; Jabbar & Nayak, 1987). Yet, despite this striking conservation of a critical cellular function, signal sequences display a remarkable lack of primary sequence homology, even among closely related proteins. This perspective first briefly reviews present understanding of signal sequence functions and then discusses results of several approaches that may enhance our understanding of the way these intriguing sequences perform their functions.

Interest in signal sequences is high. In addition to the practical motivation of finding more effective vehicles for production of proteins in recombinant systems, a better understanding of how signal sequences work will shed light on several pressing biological, biophysical, and biochemical questions. Signal sequences are essential for the efficient and selective targeting of nascent protein chains either to the endoplasmic reticulum, in eukaryotes, or to the cytoplasmic membrane, in prokaryotes. As such, they are representative of a much broader class of targeting sequences that serve as organizers and zip codes for cellular traffic of macromolecules (Warren, 1987). Furthermore, signal sequences play a central, although poorly understood, role in the translocation of polypeptide chains across membranes.

The ability of signal sequences to facilitate these complex processes despite their high degree of sequence variability (Perlman & Halvorson, 1983; Watson, 1984; von Heijne, 1985) pointedly raises the issue of the relationship between amino acid sequence and the conformations and interactions of a polypeptide chain (the so-called second half of the genetic code). Furthermore, while the importance of amino acid sequence in determining the three-dimensional structure of a mature protein has been recognized and actively investigated for the last two decades, much less attention has been devoted to the *process* of protein folding in vivo (Tsou, 1988). The sequences of existing proteins have been selected through evolution not only to adopt a functional three-dimensional structure after folding but also to optimize the protein folding process both *temporally* and *spatially*, given the constraints of the cellular context. Clearly, presence of the signal sequence (or other transient sequences) may influence the folding of the nascent chain (Park et al., 1988), and many recent results emphasize the coupling of folding and targeting (Randall & Hardy, 1986; Eilers & Schatz, 1988).

Roles and Interactions of Signal Sequences

In both prokaryotes and eukaryotes, considerable progress has been made in the last few years in the identification of components of the export or secretion machinery. However, current understanding stops abruptly at perhaps the most interesting stage of protein export: translocation across the membrane, be it cytoplasmic or ER.[1] The components and

[†] Supported by grants from the NIH (GM34962), the NSF (DCB-8896144), and the Robert A. Welch Foundation.

[‡] Dedicated to the late E. Thomas Kaiser.

[1] Abbreviations: ER, endoplasmic reticulum; SRP, signal recognition particle; SSR, signal sequence receptor; SDS, sodium dodecyl sulfate; MBP, maltose-binding protein; LPP, lipoprotein; PhoA, alkaline phosphatase; PhoE, phosphate limitation protein; PTH, parathyroid hormone; FT-IR, Fourier transform infrared; BIP, heavy chain binding protein.

0006-2960/89/0428-0923$01.50/0

mechanistic steps worked out so far are involved in *targeting* the nascent chain to the membrane and then in *cleaving* the transient signal sequence from the mature chain. Any discussion of the intermediate stages falls necessarily in the realm of speculation. Despite these gaps in our understanding, there are several points at which the signal sequence clearly must play direct or indirect roles. We consider now what is known about prokaryotic and eukaryotic protein export in light of the involvement of the signal sequence.

In higher eukaryotes, where the components of the secretory apparatus have been more fully characterized (Rapoport, 1986; Walter & Lingappa, 1986), the first interaction of the signal sequence appears to be with the signal recognition particle (SRP). This interaction is probably the first committed step in protein secretion; it ensures, by virtue of the subsequent specific binding between SRP and its receptor in the ER membrane (SRP receptor or docking protein), that the nascent chain will be correctly targeted. Under some experimental conditions, SRP binding leads to an arrest or a pause in translation (Walter & Blobel, 1981b), which is relieved by release of SRP upon its binding to SRP receptor (Meyer et al., 1982; Gilmore et al., 1982; Gilmore & Blobel, 1982). This arrest or pause may or may not be a feature of the in vivo process (Meyer, 1985); if so, it would couple synthesis to translocation by preventing translation unless delivery to the membrane had taken place.

Interaction of the signal sequence with SRP has been probed by cross-linking experiments with a photoactivatable probe. These experiments indicated that the 54-kDa subunit of the SRP ribonucleoprotein complex was the site of signal sequence binding (Kurzchalia et al., 1986; Krieg et al., 1986). This binding step occurs after the signal sequence emerges from the ribosome, viz., when a chain of about 70–80 residues has been synthesized (Wiedmann et al., 1987a). All evidence points to the existence of *only one* SRP in a particular organism; hence, several different signal sequences must be recognized by the same SRP. The binding site for signal sequences may include the ribosome, since a ternary complex of SRP, nascent chain, and ribosome forms prior to membrane targeting (Walter et al., 1981). Furthermore, signal sequence binding may cause conformational changes within SRP, since the SRP/ribosome affinity increases by 4 orders of magnitude in the presence of nascent chain (Walter & Blobel, 1981a).

The involvement of the signal sequence in the next steps, viz., association with the membrane and translocation, is unclear. Cross-linking studies analogous to those used to identify the SRP-signal sequence interaction have revealed the presence of a 35-kDa species in the ER membrane that is proposed to serve as a "signal sequence receptor" (SSR) (Wiedmann et al., 1987b). Cross-linking of isolated signal peptides identified a 45-kDa species in the microsomal membrane (Robinson et al., 1987). There may in fact be multiple interactions of the nascent chain at the level of the membrane, including the possibility of binding to phospholipids, which has often been suggested on the basis of the hydrophobicity of signal sequences (see below; von Heijne & Blomberg, 1979; Engelman & Steitz, 1983; Briggs et al., 1986). Initial binding (either to a proteinaceous receptor or to the bilayer) may be followed by interaction with a protein or complex that facilitates translocation. Since nothing is known about this process, one can speculate freely. Signal peptidase recognition and cleavage constitute the final steps in eukaryotic protein secretion that involve the signal sequence; these processes occur on the lumenal side of the ER membrane. Evans et al. (1986) isolated signal peptidase from canine pancreas as a complex of six polypeptide chains. Signal peptidase has more recently been purified from hen oviduct in a solubilized form requiring only two polypeptide chains (Baker & Lively, 1987). The multiple components of the canine microsomal peptidase complex are potential candidates for an apparatus to translocate the polypeptide chain.

The steps in prokaryotic protein export are less well defined (Randall et al., 1987), but recent findings promise clarification of the mechanism in the very near future. As is true presently for the eukaryotic systems, virtually nothing is known about translocation; the components identified thus far are nearly all involved in target or cleavage. Much of current knowledge came originally from genetic evidence [for reviews, see Bankaitis et al. (1985), Benson et al. (1985), and Oliver (1985)], which implicated the products of several genes in bacterial protein export: SecA,[2] PrlA (also known as SecY), SecB, PrlD, and the two signal peptidases, leader peptidase (or signal peptidase I) and signal peptidase II, which processes lipoproteins. Biochemical evidence has led to the identification of other species, including soluble factors that are required for in vitro translocation (Mueller & Blobel, 1984; Weng et al., 1988), and a protein, called trigger factor, which forms a complex with the precursor to OmpA and stabilizes this precursor in a translocation-competent form in vitro (Crooke & Wickner, 1987; Crooke et al., 1988a,b).

Putting together all available data at the present time, using a eukaryotic paradigm, suggests the following steps: Upon translation of the nascent protein, part or all of the precursor protein binds to cytoplasmic factors that may include SecA, trigger factor, and/or SecB, depending on the protein to be exported and the kinetic relationship of translation and translocation.[3] SecB (Collier et al., 1988; Kumamoto & Gannon, 1988) and trigger factor (Crooke & Wickner, 1987; Crooke et al., 1988a,b; Lill et al., 1988) may be most critical to export in cases where the synthesis of the precursor is complete or nearly complete prior to its entry into the export pathway. These proteins seem to be important in maintaining an export-competent conformation in the precursor; also, only a subset of proteins depends on SecB for export. The binding of precursors to SecB does not seem to require interaction with the signal sequence (Collier et al., 1988); it is not known whether binding to trigger factor does.

SecA is known to be an essential player in bacterial protein export and likely serves a role similar to that of eukaryotic SRP. Defects in SecA cause pleiotropic effects on protein export (Oliver & Beckwith, 1981). This protein has recently been purified and its gene sequenced; it has 901 amino acids and no apparent homology with any known protein (Schmidt et al., 1988). A complex of nascent chain and SecA (possibly plus trigger factor) may, by analogy with SRP/docking protein, bind to the cytoplasmic membrane and facilitate targeting of the nascent chain to export sites. Purified SecA can be added back to membranes depleted or defective in SecA

[2] The genes associated with protein export were named for this putative function, hence *sec* for secretion or *prl* for protein localization. Then different genes were named *A*, *B*, etc. The products of these various genes are designated SecA, PrlA, etc.

[3] Results of different experiments [recently reviewed by Lee and Beckwith (1986)] have been interpreted to indicate cotranslational translocation [for example, Smith et al. (1977)], posttranslational translocation (Koshland & Botstein, 1982), or domain-by-domain cotranslational translocation mechanisms (Randall, 1983). It seems increasingly clear that the relative rates of translation and of export vary as a function of the nature of the protein (for example, its size and rapidity of folding) and of the cellular conditions (i.e., whether there is high export activity and consequent saturation of export sites).

product in vitro and will reconstitute protein translocation (Cabelli et al., 1988).

PrlA (SecY) is a membrane protein (Akiyama & Ito, 1985); mutations in the *prlA* gene cause pleiotropic effects on protein export (Ito et al., 1983). PrlA has been shown to be essential in protein export both in vivo (Ito et al., 1984) and in vitro (Fendl & Tai, 1987). PrlA may be a receptor for the signal sequence (analogous to SSR), may play a direct role in translocation (as a pore or tunnel), or may serve as a receptor for the SecA/export complex (like docking protein). The first suggestion is supported by the finding that several signal sequence mutations normally associated with severe export defects are suppressed very effectively by mutations within PrlA (Emr et al., 1981). As pointed out by Randall et al. (1987), this argument is not unequivocal, since indirect effects cannot be ruled out. For example, the PrlA mutations may alter binding to another species such that its interaction with signal sequences becomes less restrictive. Nonetheless, inspection of the types of signal sequence mutations that can be tolerated in different PrlA backgrounds, and of the changes in the PrlA sequence itself, is of interest in efforts to relate the required sequence characteristics of signal peptides to their ability to function (see below). The suggestion that SecY (PrlA) is a receptor for a SecA/export complex is supported by the recent finding that purified SecA can suppress a temperature-sensitive SecY defect in translocation activity in membrane vesicles (Fandl et al., 1988).

Possible mechanisms for the translocation steps in prokaryotic protein export are, as in the eukaryotic case, speculative. The possibility of direct interaction between the nascent chain and membrane lipids has been discussed frequently (von Heijne & Blomberg, 1979; Engelman & Steitz, 1983; Briggs et al., 1986) but again lacks direct evidence. The last step involving the signal sequence is recognition and cleavage by the leader or signal peptidase. The active site of the transmembrane leader peptidase (signal peptidase I) is situated on the periplasmic side of the membrane (Zimmermann et al., 1982), requiring that the signal sequence cleavage site be oriented appropriately.

While the specific components of the eukaryotic and prokaryotic export machinery are not the same, one conclusion applies to both: The signal sequence is required to perform several roles which probably involve interactions with a variety of species. From the above discussion, we can extract a list of possible roles and interactions of signal sequences:

(*1*) *Binding to SRP or Prokaryotic Equivalent.* In the eukaryotic system, identification of a nascent chain as a secretory protein is mediated by the signal sequence/SRP interaction. Delivery of the nascent chain to the ER membrane is catalyzed by the SRP/SRP receptor (docking protein) binding step. In prokaryotes, these steps may involve SecA, trigger factor, SecB, and/or other cytoplasmic factors. SecA is a good candidate for facilitating membrane targeting of the export complex, since evidence supports its association with the membrane (Oliver & Beckwith, 1982), possibly via PrlA (SecY) (Ryan & Bassford, 1985;[4] Fandl et al., 1988). The identification of SecA mutations that suppress signal sequence defects (Ryan & Bassford, 1985;[4] Fikes & Bassford, 1989) suggests that SecA may interact directly with the signal sequence, but alternative explanations cannot be excluded. Trigger factor, on the other hand, has been found to associate with ribosomes and to interact in a saturable way with membrane vesicles in in vitro translocation assays (Lill et al., 1988). It is possible that the multiple functions of SRP, which in eukaryotes are carried out by different polypeptide chains in one ribonucleoprotein complex (Siegel & Walter, 1988a,b), are associated with separate species in prokaryotes.

(*2*) *Binding to the Membrane To Be Translocated.* This role may be mediated by a proteinaceous receptor molecule (SSR or PrlA) or by direct association with membrane lipids, or possibly both.

(*3*) *Facilitation of Translocation.* The greatest mystery of protein secretion at present is the mechanism of translocation across the membrane. The signal sequence is present at the time of initiation of translocation but may be cleaved during the transfer of the mature portion of the nascent chain. Hence, its potential role in this process might be to facilitate initiation of translocation.

(*4*) *Recognition by Signal Peptidase.* One of the most clear-cut requirements of all cleaved signal sequences is that they be recognized and productively bound by the processing enzyme. This step may involve "traditional" enzyme/substrate interactions but also is likely to be influenced by the topology of the translocating chain in the membrane. The signal sequence must be compatible with the arrangement of the peptidase and nascent chain that enables cleavage to take place (e.g., depth in the membrane, specific conformational features, interaction with the mature segment).

What Makes a Sequence Function as a Signal Sequence?

The traditional approaches to determining sequence/function correlations are quickly stymied by signal sequences. Comparison of all known signal sequences reveals no regions of strict homology: the cleavage site shows the strongest conservation, as might be expected since it must be recognized by signal peptidase. Although other portions of signal sequences lack homology, they do display common distributions of residue type. Von Heijne (1985) has shown by detailed analyses of known signal sequences that their variability is limited: Three recognizable regions with specific characteristics emerge from his comparisons. These characteristics are shared by signal sequences from both eukaryotes and prokaryotes. Counting from the cleavage site, there are usually five to seven residues [including the "−1, −3 rule" residues (von Heijne, 1983; Perlman & Halvorson, 1983)] that comprise the so-called c-region. Although not generally charged, these residues are of higher polarity on average than those in the "h-region" immediately N-terminal to the c-region. The h-region is rich in Leu, Ala, Met, Val, Ile, Phe, and Trp but may contain an occasional Pro, Gly, Ser, or Thr residue. This hydrophobic core (h-region) is the true hallmark of signal sequences. Its length (10 ± 3) distinguishes it from membrane-spanning sequences (24 ± 2 residues long) and from hydrophobic segments of globular proteins (6–8 residues in length) (G. von Heijne, personal communication). Statistical results suggest that overall hydrophobicity is the major requirement in the h-region (von Heijne, 1985). The n-region is of highly variable length and composition, but always carries a net positive charge (on average +1.7). In eukaryotes, this charge is contributed by the N-terminus and any charged residues; in prokaryotes, the N-terminus retains a formyl-Met, and the charge comes exclusively from basic residues.

This sort of analysis of signal sequences convinces one that they indeed have defining characteristics. However, relating

[4] In the paper by Ryan and Bassford (1985), the SecA mutation was referred to as PrlD2. Subsequent to sequencing, it was found to be in the SecA gene (Fikes & Bassford, 1989). This allele has an effect on export of MBP with a defective signal sequence that is synergistic with mutations in PrlA (SecY), arguing for an interaction between these PrlA and SecA.

these characteristics to the functional roles listed above is difficult. As pointed out by von Heijne (1985), the lack of specific patterns in signal sequences does not seem consistent with specific protein/protein interactions: "Both regions (n and h) seem well suited for binding in a rather unspecific way to the surface (n-region) and to the interior (h-region) of membranes."

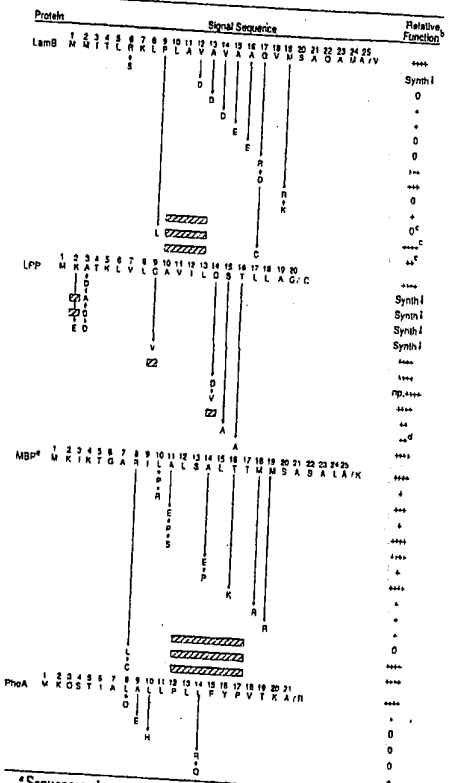
On the other hand, there have been many reports of alterations in signal sequences, including point mutations, that lead to loss of function. In fact, examination of these sequences and the specific nature of their export defects is a promising route to determining sequence/function correlations. An early example of this strategy was the incorporation of β-hydroxyleucine in the preprolactin signal sequence, which led to a cytoplasmic protein that escaped SRP binding (Hortin & Boime, 1980; Walter et al., 1981; Walter & Blobel, 1981a). Substitution of this polar Leu analogue in a nascent protein whose signal sequence has no or few Leu residues did not impair export. This result suggests that the hydrophobic core of eukaryotic signal sequences mediates their recognition by SRP. There is a plethora of data on mutations in bacterial signal sequences that impair export to varying degrees, and in some cases in quite distinct ways (Benson et al., 1985). Examples drawn from four *Escherichia coli* proteins are gathered in Table I. Two of these proteins (LamB, the λ phage receptor, and LPP, the major lipoprotein) are in the outer membrane, and two (MBP, maltose-binding protein, and PhoA, alkaline phosphatase) are periplasmic. Most of these mutations lead to accumulation of precursor in the cytoplasm. Examples of point mutations in the n-region that lead to decreased synthesis (translation) of the exported protein have been found in both lipoprotein (Inouye et al., 1982; Vlasuk et al., 1983) and LamB (Hall et al., 1983). In the case of the LamB Arg6 → Scr mutation, evidence has been presented that the synthesis-down phenotype was not a consequence of mRNA structure or stability (Benson et al., 1987) and therefore argues for a coupling of export and synthesis, as has been shown in eukaryotes under certain conditions. Suppressors of the translation-down phenotype were found and arose from incorporation of a hydrophilic residue in the h-region or deletion of a large segment of the signal sequence. One interpretation of this result is that the first mutation prevents release of an SRP-like block of translation and the second mutation bypasses this block altogether by disrupting the binding site for the SRP-like species.

The bulk of the export-defective mutants have suffered alterations in the hydrophobic core, usually introduction of a charged amino acid or sometimes a deletion. Also shown in this table are some pseudorevertants which indicate the nature of compensating changes that can again yield a functional signal sequence. In most cases, the changes leading to a reversion phenotype restore the hydrophobic core. Generally, the introduction of a charge in the h-region has a major effect regardless of the specific position; an exception was found in LamB where a charge at position 17 is only modestly deleterious, but a charge at position 19 nearly abolishes export.

On the one hand, all of these mutations demonstrate the sensitivity of signal sequences to quite modest changes. On the other hand, it seems as though debilitating mutations are very rare: Even fairly substantial alterations in the signal sequence usually show somewhat "leaky" phenotypes (Ferenci & Silhavy, 1987). It may be that the numerous cytoplasmic factors (SecB, trigger, SecA, etc.) can rescue the cell from accumulation of precursors in the case of weak signal sequences. Furthermore, a point mutation in PrlA (the PrlA4 allele) suppresses several severe mutations of signal sequences (Stader et al., 1986). While the mechanism of this suppression is a puzzle,[5] it is tempting to speculate that PrlA forms a pore or channel that becomes more permissive in the PrlA4 strains.

In light of the variability of wild-type signal sequences and the apparent tolerance to many mutational variations, one might well ask the question "Are there sequences that will *not* work as signal sequences?" This question was addressed by Kaiser et al. (1987), who substituted random sequences for the signal sequence of yeast invertase and asked for secretion. Their assay for function was relative growth on sucrose, and they found remarkably that 20% of random sequences from a human DNA library would work. Not surprisingly, it is difficult to score *functional* versus *nonfunctional* in a clear-cut way. The measure used by these authors did not always correlate with near wild-type levels of secreted invertase. Nonetheless, those sequences that facilitated invertase export at reasonable levels had the characteristics of signal sequences as described above. Hence, a result that initially seemed to point a pronounced lack of constraints on signal sequences actually reconfirms that we have an idea of what defines a signal sequence. Furthermore, as pointed out by Ferenci and Silhavy (1987), known signal sequences have been optimized for the particular passenger protein and the needs of the cell. Their ability to function in vivo in most cases will go well beyond meeting some minimal level of export. The multiple roles played by signal sequences and the likelihood of additional mechanisms for facilitating or "rescuing" export (see above) also confound the interpretation of the invertase/random sequence results.

Table I: Examples of Signal Sequence Mutations Causing Export Defects in *E. coli*[a]



[a] Sequences shown were compiled by Benson et al. (1985); original references can be found there except as noted. Hatched boxes represent deletions; vertical arrows indicate point mutations. Mutant sequences are otherwise unchanged from wild type. [b] Wild-type levels of export are designated and those severely defective by 0. Intermediate levels of export are then qualitatively indicated by +++, ++, or +. Reduced levels of synthesis are noted (Synth I). np indicates not processed. [c] Emr and Silhavy (1983). [d] This mutant shows slow processing to mature form. [e] Bankaitis et al. (1985).

Another approach to determining the limits on signal sequences is to idealize them and ask whether the assumptions used in the idealization were justified. Kendall et al. (1986) applied this approach to the alkaline phosphatase signal sequence and have been able to replace the entire hydrophobic core by Leu, or more recently by Ile (Kendall & Kaiser, 1988), while retaining function. In a similar study, the h-region of the hen lysozyme signal sequence was replaced by $(Leu)_m$, and the amount of mature lysozyme secreted to the medium by *Saccharomyces cerevisiae* was determined. Best export occurred with a core length of 8–10 residues (Yamamoto et al., 1987).

Emerging from all of these approaches is the generalization that primary structure is not critical to signal sequence functions. Clearly, disruption of the hydrophobic core leads to a less effective signal sequence. In all cases, it is difficult to deduce the point in the export pathway where a defect is manifesting itself. In vitro translocation assays may help to sort out the steps, as may genetic tests for suppression.

The Study of Isolated Signal Peptides

We and others have sought a better understanding of how signal sequences work by studying them as isolated peptides. We can then analyze their conformations, their interactions, and their responses to changes in environment. This dissection strategy can be risky, but the characteristics of signal sequences reviewed in the previous section seem to invite such an approach. The fact that they can in many instances be transferred from one protein to another and still function implies that they act quite independently of their context (the sequences adjacent). Signal sequences perform their multiple roles while they are attached as N-terminal extensions on their cognate mature proteins; yet they are probably relatively free of interactions with the rest of the nascent chain.[6] Signal sequences seem likely to interact with many cellular components, some of which have been characterized (signal recognition particle, signal peptidase) and others postulated ["translocon" (Walter & Lingappa, 1986), membrane lipids, signal sequence receptor], but they apparently do so by virtue of their overall properties (residue type and patterns of residues) as opposed to specific sequence. Characterization of isolated signal sequences has the potential to reveal what the critical properties are, particularly if *functional* signal sequences are compared to variants that are *nonfunctional*. Results to date using this strategy have been enlightening.

A potential limitation in the study of isolated signal sequences is that they are not likely to have strongly preferred conformations. Linear peptides of fewer than 30 residues generally sample several conformations in aqueous solution (Wright et al., 1988). Even in structure-promoting environments, most short polypeptides are likely to be interconverting among different structures, with at best a bias toward one. Characterization of such a dynamic state is extremely difficult. On the other hand, the biological roles of signal sequences may require them to be conformationally dynamic and to respond to different environments by conformational changes. Studies of isolated signal peptides suggest this to be the case.

Signal sequences as isolated peptides have generally demonstrated similar conformational preferences (Table II), which had been predicted from secondary structure analysis (Austen, 1979; Rosenblatt et al., 1980; Emr & Silhavy, 1983).[7] For the most part they are unordered in aqueous solution, and interactions with nonpolar solvents or with micellar solutions induce adoption of α-helix. Preproparathyroid hormone (signal sequence plus six residues of the pro region plus one residue of the mature hormone) was found to exist predominantly in β-structure in aqueous solution (Rosenblatt et al., 1980), and other examples have been reported of a conformational equilibrium that includes β forms (Batenburg et al., 1988). It is clear that conformational interconvertibility is a feature of signal sequences, and it has been suggested to be of functional importance (Austen & Ridd, 1981; Bedouelle & Hofnung, 1981; Rosenblatt et al., 1980; Batenburg et al., 1988a). As a complicating factor, isolated signal peptides are sparingly soluble, and one must interpret with caution the presence of β-structure. We have followed the circular dichroism spectra

Table II: Signal Sequences Studied as Isolated Peptides

peptide	sequence	conformation[a] aqueous	conformation[a] nonpolar[b]	reference
pre-pro-PTH[c]	SAKDMVKVMIVMLAICFLARSDGKSVKKR(Y)	β	α	Rosenblatt et al. (1980)
M13 coat protein	MKKSLVLKASVAVATLVPMLSFA-NH_2	rc	α	Shinnar and Kaiser (1984)
modified pretrypsinogen	Ac-NPKKAKLFLFLALLLAYVA	rc	α	Austen and Ridd (1982)
PhoE	MKKSTLALVVMGIVASASVQA	β	α	Batenburg et al. (1988a,b)
LamB (and several mutants)	MMITLRKLPLAVAVAAGVMSAQAMA	rc	α	Briggs (1986)
OmpA	MKKTAIAIAVALAGFATVAQA/APKD	rc/β[d]	α	D. W. Hoyt, unpublished results

[a] Predominant conformation from CD analysis; rc designates random or unordered structure. [b] Nonpolar environments include trifluoroethanol, SDS micelles, or hexafluoroisopropyl alcohol. [c] The underlined residues are from the pro region of the hormone; the C-terminal Tyr residue was D. [d] As noted in the text, this peptide undergoes a time-dependent conformational change from a random ensemble of states to β-structure.

[5] The PrlA4 allele arises from a Leu → Asn substitution (Stader et al., 1986) in the last predicted transmembrane-spanning helix of the protein (Akiyama & Ito, 1987). The predicted structure has 10 membrane-spanning segments, which are consistent with alkaline phosphatase topology mapping (Akiyama & Ito, 1987).

[6] Evidence in favor of this idea includes the observation that signal sequences are accessible to antibodies (Baty & Lazdunski, 1979) and that they can be proteolytically clipped from the precursor species and will bind nonionic detergents while they are still linked to the precursor (Dierstein & Wickner, 1985). Evidence against this image includes the recent report that the presence of the signal sequence on maltose-binding protein modulates its rate of folding in vitro (Park et al., 1988), which suggests direct interaction between the signal sequence and the mature region of the protein.

[7] It is quite surprising that these predictions would apply to signal sequences, since they have been derived from the behavior of sequences within globular proteins.

of aqueous solutions of the OmpA signal sequence from *E. coli* as a function of time; this peptide begins in an unordered conformational ensemble and gradually changes to nearly 100% β (David W. Hoyt, unpublished results). The rate of this conformational transition is increased by higher concentration and is decreased at low pH. Intermolecular association is the apparent driving force for the conformational change. Nonetheless, the fact that these sequences visit both α-helical and β-structures argues that these states are of very similar energy.

Assessing the importance of these preferred conformations of isolated signal peptides in terms of their function in vivo is not straightforward. It is difficult to mimic the microenvironments likely to be encountered in the export process, and it is not clear whether a particular conformational propensity is required for function. To address these problems, we have made use of the families of export-impaired mutant signal sequences from *E. coli* to draw correlations between physical properties and ability to facilitate export in vivo. The LamB system was chosen and offers several particularly interesting comparisons: For example, as shown in Table I, a deletion of four residues in the h-region of the LamB sequence causes a severe export defect. This is not surprising given the generality of the requirement for a 10–12 residue hydrophobic core. What is surprising is that two pseudorevertant strains with restored ability to facilitate export were isolated from the deletion mutant strain; the pseudorevertants had secondary point mutations that apparently compensate for the loss of four residues (Emr & Silhavy, 1983). When Emr and Silhavy found these strains, they argued that α-helicity is required for signal sequence function, since the deletion mutant would be predicted (Chou & Fasman, 1974a,b) to have a much reduced tendency to adopt helix (relative to wild type) because of the proximity of a Pro and a Gly in its sequence. The two pseudorevertants replace either the Pro or the Gly with a helix-favoring residue and hence restore predicted helix formation. Conformational analysis of these sequences as isolated peptides confirms this interpretation (Briggs & Gierasch, 1984; Briggs, 1986). We find that the wild-type LamB signal sequence adopts a largely α-helical conformation in SDS micellar environments, in lipid vesicles, or in water/trifluoroethanol mixtures. The deletion mutant has much less helix under the same conditions, and the pseudorevertants show increased helicity.

The ability to take up an α-helix in nonpolar or interfacial environments thus seems to be a property of functional signal sequences, but it is clearly not sufficient for a given sequence to function as a signal sequence. For example, we have also examined the two LamB signal sequence mutants that harbor a charge (A13D and G17R,[1] Table I) as isolated peptides (C. J. McKnight, M. S. Briggs, and L. M. Gierasch, unpublished results). Although the extent to which they cause an export-defective phenotype in vivo is quite different, their tendency to adopt α-helix is not; both behave similarly to wild type.

As noted above, the possibility that signal sequences interact with the membrane has been suggested on many occasions. Isolated signal peptides provide a means of exploring the likelihood and mechanism of such an interaction. Furthermore, comparison of the various mutant signal sequences confirms that a high affinity for a phospholipid membrane is also characteristic of functional signal sequences. We have compared the abilities of the various LamB mutant signal sequences to insert either into a lipid monolayer or into a lipid bilayer in a vesicle (Briggs et al., 1985; C. J. McKnight, M. S. Briggs, and L. M. Gierasch, unpublished results). We found the wild type, the G17R, and the Pro → Leu pseudorevertant to have the highest affinities. The A13D mutant, which is severely export defective in vivo yet folds into helix equally as well as the G17R, has a reduced affinity for a membrane. Others have reported high-affinity lipid interactions for signal sequences from M13 (Shinnar & Kaiser, 1984), from PhoE (Batenburg et al., 1988b), and from ovalbumin (Fidelio et al., 1987).

In order to describe more fully the conformational states of the LamB wild-type signal sequence upon its interaction with a membrane, we have carried out spectroscopy on peptide/lipid monolayers transferred onto solid supports [either quartz plates for CD or germanium crystals for Fourier transform infrared (FT-IR) spectroscopy] (Briggs et al., 1986). The transfer was done under two conditions: either at a high packing density (surface pressure) of the lipid, such that the signal peptide did not insert but instead associated with the surface, or at a lower lipid packing density (surface pressure resembling that of a biological membrane), such that the signal peptide inserted into the lipid acyl chain region. We found that the peptide adopted a β-structure when associated with the surface but was predominantly α-helical when inserted. From differential scanning calorimetry (M. Kodama, M. S. Briggs, C. J. McKnight, L. M. Gierasch, and E. Freire, unpublished results), fluorescence studies of Trp-containing signal peptides (C. J. McKnight and M. Rafalski, unpublished results), and polarized FT-IR (D. G. Cornell, R. A. Dluhy, C. J. McKnight, and L. M. Gierasch, unpublished results), we have concluded that α-helical, inserted form of the LamB wild-type signal peptide is oriented parallel to the acyl chains. Assuming that the N-terminus does not traverse the membrane, this mode of interaction suggests that an isolated signal peptide can facilitate the insertion and translocation of its C-terminus to the opposite side of the membrane. We have incorporated this idea and the associated conformational interconversions into a model for the initial interactions of the signal sequence with a membrane in protein export (Briggs et al., 1986). We have now synthesized the LamB wild-type signal sequence plus a segment of the mature protein in order to ask whether the signal sequence can cause the C-terminal segment to be translocated in the absence of any other components of the export apparatus.

These observations on isolated signal sequences serve to point out just what a functional signal sequence will do, by virtue of its inherent properties that arise from its amino acid sequence. Yet, there is no question that protein export in vivo involves additional components and that the signal sequence interacts with proteins that target and possibly translocate the bulk of the nascent chain. In fact, isolated signal peptides can be used as probes of the export machinery. As noted above, Robinson et al. (1987) used this approach with a photolabile cross-linker on the signal peptide to find a possible component of the ER translocation apparatus. Addition of synthetic signal peptides at approximately micromolar concentration to in vitro translocation systems causes inhibition of translocation both in prokaryotes (Chen et al., 1987) and in eukaryotes (Majzoub et al., 1980; Austen & Ridd, 1983; Austen et al., 1984). The LamB mutant signal sequences inhibited the translocation of pre-alkaline phosphatase and pre-OmpA to an extent that paralleled their in vivo function (Chen et al., 1987). This result supports the interpretation that the inhibition arises from an intervention of the added signal peptide at a normal step in export, despite the relatively high concentrations required. However, we could not distinguish a mode of inhibition based on competition between the signal peptide and the precursor for a proteinaceous receptor (cytoplasmic or membrane associated) from one based on membrane insertion and an indirect effect on translocation. Recently, we found that an all-D LamB wild-type signal peptide inhibits translocation of the pre-OmpA less than does the all-L peptide, arguing that there is a recognition by protein, which would distinguish the opposite handedness of the all-D peptide (A. R. Sgrignoli, L. L. Chen, P. C. Tai, and L. M. Gierasch, unpublished results).

[1] Substitution mutations are designated by the single-letter code for the original residue, the position (numbered from the N-terminus), and then the single-letter code for the substituted residue: hence, A13D, etc.

Conclusions: Implications for Signal Sequence Function

Signal sequences mediate a critical cellular function: correct and efficient localization of nascent secretory proteins. Yet, paradoxically their amino acid sequences are not highly constrained. As discussed in this perspective, they must interact with several components of the export pathway, whether in prokaryotes or in eukaryotes. These interactions are intriguing in their lack of a requirement for specific sequences. Similar binding mechanisms may be operative in other systems: for example, in presentation of antigens by the major histocompatibility complexes (Bjorkman et al., 1987), in binding to BiP in the ER lumen (Gething et al., 1986), and in degradative proteolysis as mediated by protease La in *E. coli* (Waxman & Goldberg, 1986). In all of these examples, as in the case of signal sequences, the overall properties of sequences are the key recognition features. In addition, the way signal sequences are presented probably contributes to their ability to facilitate the correct targeting of a nascent chain despite their lack of sequence specificity. Since they are on or near the N-terminus and accessible (not sequestered by folding), it is likely that SRP or its prokaryotic equivalent binds to the growing polypeptide chain and specifies targeting to the ER or cytoplasmic membrane whenever a signal-sequence-like pattern of residues emerges early in translation in a largely unfolded form. As demonstrated by Kaiser et al. (1987), many sequences within a mature polypeptide *could* function as signal sequences; that they do not is probably a consequence of their mode of presentation and their relationship to the three-dimensional structure of the protein. It could be said that cytoplasmic proteins have to be selected *not* to reveal any targeting sequences so as not to be incorrectly localized. Perhaps more rapid folding is required of nascent chains destined to remain in the cytoplasm.

Because of these characteristics of signal sequences—that they function by virtue of their overall properties and quite independently of their context, work on isolated signal peptides has been particularly fruitful. Signal sequences are clearly conformationally flexible, responding to their environment by pronounced conformational changes. They also have a strong tendency to insert into phospholipid membranes. This biophysical attribute may have a direct functional significance, implying interactions with lipids in vivo. Alternatively, the binding sites for signal sequences on the various proteinaceous components of the export pathway may require the same linear amphiphilicity that favors lipid interactions, perhaps because at any earlier stage in evolution there were direct lipid interactions. Further understanding of these questions awaits dissection of the components required for export and analysis of their interactions with signal sequences.

Acknowledgments

The work from my laboratory has been carried out through the efforts of a number of people: Marty Briggs, Jamie McKnight, Dave Hoyt, Maria Rafalski, Anita Sgrignoli, Robin Huff, and Jennifer Johnston. Fruitful collaborations with Don Cornell, Rich Dluhy, Tom Silhavy, P. C. Tai, Masayori Inouye, and Ernesto Freire have also made possible several aspects of the work. I thank the several colleagues who provided results prior to publication.

References

Akiyama, Y., & Ito, K. (1985) *EMBO J. 4*, 3351–3356.
Akiyama, Y., & Ito, K. (1987) *EMBO J. 6*, 3465–3470.
Austen, B. A. (1979) *FEBS Lett. 103*, 308–313.
Austen, B. A., & Ridd, D. H. (1981) *Biochem. Soc. Symp. 46*, 235–258.
Austen, B. M., & Ridd, D. H. (1983) *Biochem. Soc. Trans. 11*, 160–161.
Austen, B. M., Hermon-Taylor, J., Kaderbhai, M. A., & Ridd, D. H. (1984) *Biochem. J. 224*, 317–325.
Baker, R. K., & Lively, M. O. (1987) *Biochemistry 26*, 8561–8567.
Bankaitis, V. A., Ryan, J. P., Rasmussen, B. A., & Bassford, P. J., Jr. (1985) *Curr. Top. Membr. Transp. 54*, 105–150.
Batenburg, A. M., Brasseur, R., Ruysschaer, J.-M., van Scharrenburg, G. J. M., Slotboom, A. J., Demel, R. A., & de Kruijff, B. (1988a) *J. Biol. Chem. 263*, 4202–4207.
Batenburg, A. M., Demel, R. A., Verkleij, A. J., & de Kruijff, B. (1988b) *Biochemistry 27*, 5678–5685.
Baty, D., & Lazdunski, C. (1979) *Eur. J. Biochem. 102*, 503–507.
Bedouelle, H., & Hofnung, M. (1981) in *Membrane Transport and Neuroreceptors*, pp 399–403, Alan R. Liss, New York.
Benson, S. A., Hall, M. N., & Silhavy, T. J. (1985) *Annu. Rev. Biochem. 54*, 101–134.
Benson, S. A., Hall, M. N., & Rasmussen, B. A. (1987) *J. Bacteriol. 169*, 4686–4691.
Bird, P., Gething, M.-J., & Sambrook, J. (1987) *J. Cell Biol. 105*, 2905–2914.
Bjorkman, P., Saper, M. A., Samroui, B., Bennett, W. S., Strominger, J. L., & Wiley, D. C. (1987) *Nature 329*, 512–517.
Briggs, M. S. (1986) Ph.D. Dissertation, Yale University.
Briggs, M. S., & Gierasch, L. M. (1984) *Biochemistry 23*, 3111–3114.
Briggs, M. S., & Gierasch, L. M. (1986) *Adv. Protein Chem. 38*, 109–180.
Briggs, M. S., Gierasch, L. M., Zlotnick, A., Lear, J. D., & DeGrado, W. F. (1985) *Science 228*, 1096–1099.
Briggs, M. S., Cornell, D. G., Dluhy, R. A., & Gierasch, L. M. (1986) *Science 233*, 206–208.
Cabelli, R. J., Chen, L., Tai, P. C., & Oliver, D. B. (1988) *Cell 55*, 683–692.
Cerretti, D. P., Dean, D., Davis, G. R., Bedwell, D. M., & Nomura, M. (1983) *Nucleic Acids Res. 11*, 2599–2616.
Chen, L., Tai, P. C., Briggs, M. S., & Gierasch, L. M. (1987) *J. Biol. Chem. 262*, 1427–1429.
Chou, P. Y., & Fasman, G. D. (1974a) *Biochemistry 13*, 211–222.
Chou, P. Y., & Fasman, G. D. (1974b) *Biochemistry 13*, 222–245.
Collier, D. N., Bankaitis, V. A., Weiss, J. B., & Bassford, P. J., Jr. (1988) *Cell 53*, 273–283.
Crooke, E., & Wickner, W. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 5216–5220.
Crooke, E., Brundage, L., Rice, M., & Wickner, W. (1988a) *EMBO J. 7*, 1831–1835.
Crooke, E., Guthrie, B., Lecker, S., Lill, R., & Wickner, W. (1988b) *Cell 54*, 1003–1011.

Dierstein, R., & Wickner, W. (1985) *J. Biol. Chem. 260*, 15919–15924.
Eilers, M., & Schatz, G. (1988) *Cell 52*, 481–483.
Emr, S. D., & Silhavy, T. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4599–4603.
Emr, S. D., Hanley, S., & Silhavy, T. J. (1981) *Cell 23*, 79–88.
Engelman, D. M., & Steitz, T. A. (1981) *Cell 23*, 411–422.
Evans, E. A., Gilmore, R., & Blobel, G. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 581–585.
Fandl, J. P., & Tai, P. C. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 7048–7052.
Fandl, J. P., Cabelli, R., Oliver, D., & Tai, P. C. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8953–8957.
Ferenci, T., & Silhavy, T. J. (1987) *J. Bacteriol. 169*, 5339–5340.
Fidelio, G. D., Austen, B. M., Chapman, D., & Lucy, J. A. (1987) *Biochem. J. 244*, 295–301.
Fikes, J. D., & Bassford, P. J., Jr. (1989) *J. Bacteriol.* (in press).
Gething, M.-J., McCammon, K., & Sambrook, J. (1986) *Cell 46*, 939–950.
Gilmore, R., & Blobel, G. (1983) *Cell 35*, 677–685.
Gilmore, R., Walter, P., & Blobel, G. (1982) *J. Cell Biol. 95*, 453–469.
Hall, M. N., Gabay, J., & Schwartz, M. (1983) *EMBO J. 2*, 15–19.
Hortin, G., & Boime, I. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1356–1360.
Inouye, S., Soberon, X., Franceschini, T., Nakamura, K., Itakura, K., & Inouye, M. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 3438–3441.
Ito, K., Wittekind, M., Nomura, M., Shiba, K., Yura, T., Miura, A., & Nashimoto, H. (1983) *Cell 32*, 789–797.
Ito, K., Yura, T., & Cerretti, D. (1984) *EMBO J. 3*, 631–635.
Jabbar, M. A., & Nayak, D. P. (1987) *Mol. Cell. Biol. 7*, 1476–1485.
Kaiser, C. A., Preuss, D., Grisafi, P., & Botstein, D. (1987) *Science 235*, 312–317.
Kendall, D. A., & Kaiser, E. T. (1988), *J. Biol. Chem. 263*, 7261–7265.
Kendall, D. A., Bock, S. C., & Kaiser, E. T. (1986) *Nature 321*, 706–708.
Koshland, D., & Botstein, D. (1982) *Cell 30*, 893–902.
Kumamoto, C. A., & Gannon, P. M. (1988) *J. Biol. Chem.* (in press).
Lee, C., & Beckwith, J. (1986) *Annu. Rev. Cell Biol. 2*, 315–336.
Lill, R., Crooke, E., Guthrie, B., & Wickner, W. (1988) *Cell 54*, 1013–1018.
Majzoub, J. A., Rosenblatt, M., Fennick, B., Maunus, R., Kronenberg, H. M., Potts, J. T., Jr., & Habener, J. F. (1980) *J. Biol. Chem. 255*, 11478–11483.
Meyer, D. I. (1985) *EMBO J. 4*, 2031–2033.
Meyer, D. I., Krause, E., & Dobberstein, B. (1982) *Nature 297*, 503–508.
Mueller, M., & Blobel, G. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7737–7741.
Mueller, M., Ibrahimi, I., Chang, C. N., Walter, P., & Blobel, G. (1982) *J. Biol. Chem. 257*, 11860–11863.
Oliver, D. B., & Beckwith, J. (1981) *Cell 25*, 765–772.
Oliver, D. B., & Beckwith, J. (1982) *Cell 30*, 311–319.
Park, S., Liu, G., Topping, T. B., Cover, W. H., & Randall, L. L. (1988) *Science 239*, 1033–1035.
Perlman, D., & Halvorson, H. O. (1983) *J. Mol. Biol. 167*, 391–409.
Randall, L. L. (1983) *Cell 33*, 231–240.
Randall, L. L., & Hardy, S. J. S. (1986) *Cell 46*, 921–928.
Randall, L. L., Hardy, S. J. S., & Thom, J. R. (1987) *Annu. Rev. Microbiol. 41*, 507–541.
Rapoport, T. A. (1986) *CRC Crit. Rev. Biochem. 20*, 73–137.
Robinson, A., Kaderbhai, M. A., & Austen, B. M. (1987) *Biochem. J. 242*, 767–777.
Rosenblatt, M., Beaudette, N. V., & Fasman, G. D. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 3983–3987.
Ryan, J. P., & Bassford, P. J., Jr. (1985) *J. Biol. Chem. 260*, 14832–14837.
Schmidt, M. G., Rollo, E. E., Grodberg, J., & Oliver, D. B. (1988) *J. Bacteriol. 170*, 3404–3414.
Shinnar, A. E., & Kaiser, E. T. (1984) *J. Am. Chem. Soc. 106*, 5006–5007.
Siegel, V., & Walter, P. (1988a) *Trends Biochem. Sci. 13*, 314–315.
Siegel, V., & Walter, P. (1988b) *Cell 52*, 39–49.
Smith, W. P., Tai, P. C., Thompson, R. C., & Davis, B. D. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2830–2834.
Stader, J., Benson, S. A., & Silhavy, T. J. (1986) *J. Biol. Chem. 261*, 15075–15080.
Tsou, C. L. (1988) *Biochemistry 27*, 1809–1812.
Verner, K., & Schatz, G. (1988) *Science 241*, 1307–1313.
Vlasuk, G., Inouye, S., Ito, H., Itakura, K., & Inouye, M. (1983) *J. Biol. Chem. 258*, 7141–7148.
von Heijne, G. (1983) *Eur. J. Biochem. 133*, 17–21.
von Heijne, G. (1985) *J. Mol. Biol. 184*, 99–105.
von Heijne, G. (1986) *J. Mol. Biol. 192*, 287–290.
von Heijne, G., & Blomberg, C. (1979) *Eur. J. Biochem. 97*, 175–181.
Walter, P., & Blobel, G. (1981a) *J. Cell Biol. 91*, 551–557.
Walter, P., & Blobel, G. (1981b) *J. Cell Biol. 91*, 557–561.
Walter, P., & Lingappa, V. (1986) *Annu. Rev. Cell Biol. 2*, 499–516.
Walter, P., Gilmore, R., & Blobel, G. (1984) *Cell 38*, 5–8.
Warren, G. (1987) *Nature 327*, 17–18.
Watson, M. E. E. (1984) *Nucleic Acids Res. 12*, 5145–5164.
Waxman, L., & Goldberg, A. L. (1986) *Science 232*, 500–503.
Weng, Q. P., Chen, L. L., & Tai, P. C. (1988) *J. Bacteriol. 170*, 126–131.
Wickner, W., & Lodish, H. (1985) *Science 230*, 400–407.
Wiedmann, M., Kurzchalia, T. V. Bielka, H., & Rapoport, T. A. (1987a) *J. Cell Biol. 104*, 201–209.
Wiedmann, M., Kurzchalia, T. V., Hartmann, E., & Rapoport, T. A. (1987b) *Nature 328*, 830–833.
Wright, P. E., Dyson, H. J., & Lerner, R. A. (1988) *Biochemistry 27*, 7167–7175.
Yamamoto, Y., Taniyama, Y., Kikuchi, M., & Ikehara, M. (1987) *Biochem. Biophys. Res. Commun. 149*, 431–436.
Yost, C. S., Hedgepeth, J., & Lingappa, V. R. (1983) *Cell 34*, 759.
Zimmermann, R., Watts, C., & Wickner, W. (1982) *J. Biol. Chem. 257*, 6529–6536.

Biochemistry 1989, *28*, 931–935

Articles

Structure of Cytochrome b_5 in Solution by Fourier-Transform Infrared Spectroscopy†

Peter W. Holloway* and Henry H. Mantsch

Division of Chemistry, National Research Council of Canada, Ottawa, Ontario, Canada K1A 0R6

Received July 25, 1988; Revised Manuscript Received September 28, 1988

ABSTRACT: Fourier-transform infrared spectroscopy was used to examine the secondary structure of rabbit liver cytochrome b_5 and the polar and nonpolar domains of the protein. The data for both the polar and nonpolar domains agree well with those previously obtained by other physical techniques. In particular it was found that the nonpolar membrane-binding domain was predominantly α helix and that the polar domain was also highly helical, but not all α helix. The independence of the two domains in the whole molecule was, in general, confirmed by the additivity of the spectra of the two domains. The small differences that were seen indicate that there is a loss of α helix when the protein is cut into the two domains. In addition, there appeared to be a slight difference in the exposure to solvent of the amide NH groups in the α-helical portion of the nonpolar domain when it was examined in isolation.

Cytochrome b_5, an amphipathic integral membrane protein found in the endoplasmic reticulum, plays an important role in lipid metabolism (Holloway, 1983) and, because of its relative ease of purification (Ozols, 1974), has been a popular subject for model membrane studies. Several studies have probed the topography of the protein when it is bound to lipid vesicles via its nonpolar membrane-binding domain (Flemming et al., 1979; Takagaki et al., 1983; Gogol & Engelman, 1984; Markello et al., 1985, Kleinfeld & Lukacovic, 1985; Everett et al., 1986; Rzepecki et al., 1986; Arinc et al., 1987) but a complete interpretation of these data is hampered by the lack of information on the secondary structure of this domain. The polar heme-containing domain has been crystallized and subjected to X-ray analysis (Mathews et al., 1971), but the whole native protein, and the nonpolar domain in particular, have not yet been crystallized. Structures have been predicted for the nonpolar domain based on Chou and Fasman calculations and circular dichroism measurements (Visser et al., 1975; Dailey & Strittmatter, 1978; Tajima et al., 1978; Hlavica, 1984), but it was felt that the recent advances in FT-IR spectroscopy would be able to give an alternate approach to the determination of the secondary structure of the nonpolar domain. The results in this paper are in good agreement with the circular dichroism measurements.

Materials and Methods

Cytochrome b_5 was prepared as described previously (Markello et al., 1985). The protein was cleaved with TPCK-trypsin (in a ratio of 1 trypsin/10 cytochrome molecules) in 10 mM Tris-acetate, pH 8.1, containing 10 mM $CaCl_2$. The mixture was kept at 4 °C overnight, and the cloudy mixture was centrifuged at 10000*g* for 10 min. UV spectral analysis indicated almost complete retention of the polar domain in the supernatant with the nonpolar domain in the precipitate. The polar domain was purified further by chromatography on DEAE-cellulose with a linear gradient from 50 to 200 mM potassium phosphate buffer (pH 7.2). The nonpolar domain was dissolved in glacial acetic acid and subjected to gel filtration on Sephadex G-75 in 50% acetic acid–water. The peptide-containing fractions, which eluted just before the green heme band, were pooled and lyophilized. All proteins and peptides gave single bands on gel electrophoresis. Before FT-IR analysis, all samples were subjected to gel filtration on Sephadex G-25 in 10 mM NH_4HCO_3 to remove acetate ions (which produce an interfering infrared band at 1560 cm^{-1}) and lyophilized.

Infrared spectra were recorded at 22 °C with a Digilab FTS-60 instrument using a high-sensitivity DTGS detector. For each spectrum 256 interferograms were collected, coadded, apodized with a Bessel function, and Fourier transformed to give a resolution of 2 cm^{-1}. Samples were prepared in 50 mM HEPES buffer in D_2O (pD 8.0) at protein concentrations between 1 and 2 mM and were assembled between CaF_2 windows separated with a 50-μm Teflon spacer. For spectra in H_2O a 6-μm spacer was used. Fourier self-deconvolution was performed by using a Lorentzian of 25-cm^{-1} half-bandwidth and a resolution enhancement factor (*k* value) of 2.8.

Results and Discussion

Polar Domain. The spectra of the polar domain in H_2O and D_2O are shown in Figure 1A. Infrared spectra of proteins in H_2O typically show two broad bands, one between 1700 and 1600 cm^{-1} and one between 1600 and 1500 cm^{-1}. The former band is the amide I band, and the latter is a complex of the amide II band and bands due to side-chain vibrations. Both amide bands are complex composites of several discrete bands that are characteristic of specific types of secondary structure in the protein. In D_2O the accessible N–H groups will undergo H → D exchange and there will be a small shift of the component bands of the broad amide I band and a large shift, down to below 1500 cm^{-1}, of the amide II band. As seen in Figure

† Issued as NRCC No. 29208. Supported by Grant GM 23858 from USPHS to P.W.H.

* To whom correspondence should be addressed at the Department of Biochemistry, University of Virginia School of Medicine, Charlottesville, VA 22908.

0006-2960/89/0428-0931$01.50/0

Voet and Voet, Biochemistry, 2nd Ed., 1995, John Wiley & Sons

7-2C). The polypeptides of **procollagen** (Fig. 30-55) differ from those of the mature protein by the presence of both N-terminal and C-terminal propeptides of ~100 residues whose sequences, for the most part, are unlike those of mature collagen. The procollagen polypeptides rapidly assemble, *in vitro* as well as *in vivo*, to form a collagen triple helix. In contrast, polypeptides extracted from mature collagen will reassemble only over a period of days, if at all. *The collagen propeptides are apparently necessary for proper procollagen folding.*

FIGURE 30-55. An electron micrograph of procollagen aggregates that have been secreted into the extracellular medium. [Courtesy of Jerome Gross, Harvard Medical School.]

The N- and C-terminal propeptides of procollagen are respectively removed by **amino-** and **carboxylprocollagen peptidases** (Fig. 30-56), which may also be specific for the different collagen types. An inherited defect of aminoprocollagen peptidase in cattle and sheep results in a bizarre condition, **dermatosparaxis,** that is characterized by extremely fragile skin. An analogous disease in man, **Ehlers–Danlos syndrome VII,** is caused by a mutation in one of the procollagen polypeptides that inhibits the enzymatic removal of its aminopropeptide. Collagen molecules normally spontaneously aggregate to form collagen fibrils (Fig. 7-33 and 7-34). However, electron micrographs of dermatosparaxic skin show sparse and disorganized collagen fibrils. *The retention of collagen's aminopropeptides apparently interferes with proper fibril formation.* (The dermatosparaxis gene was bred into some cattle herds because heterozygotes produce tender meat.)

Signal Peptides Are Removed from Nascent Proteins by a Signal Peptidase

Many transmembrane proteins or proteins that are destined to be secreted are synthesized with an N-terminal **signal peptide** of 13 to 36 predominantly hydrophobic residues. According to the **signal hypothesis** (Section 11-4B), a

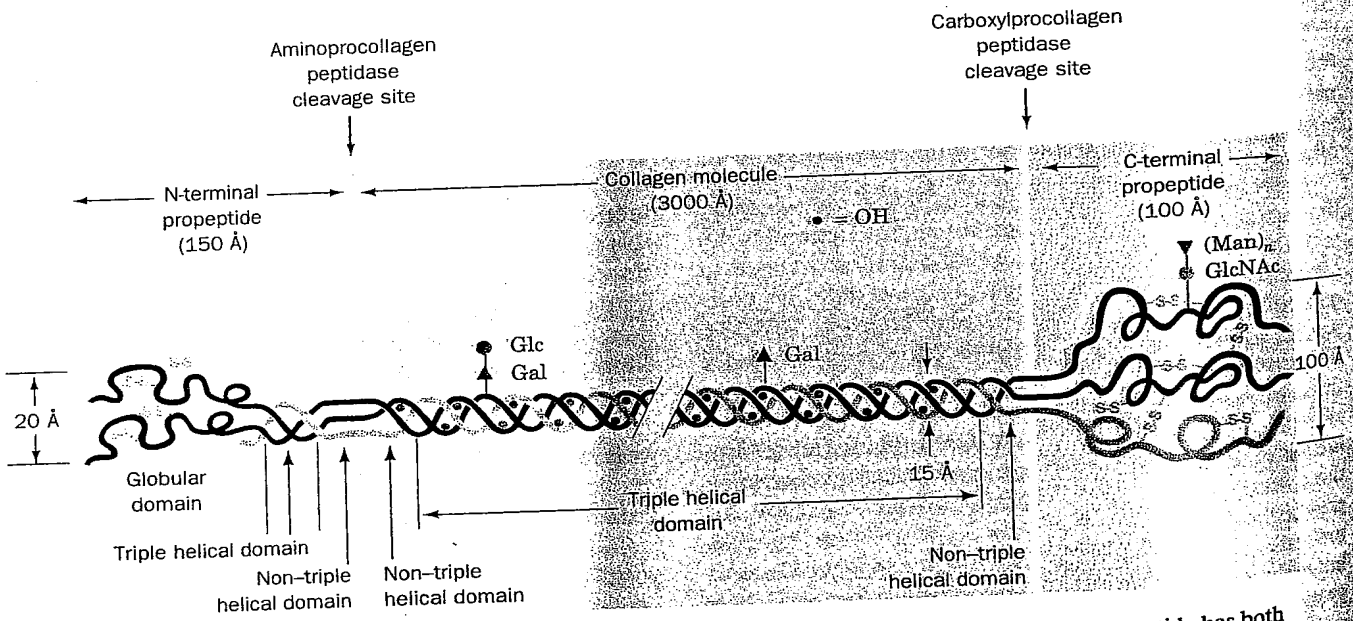


FIGURE 30-56. A schematic representation of the procollagen molecule. Gal, Glc, GlcNAc, and Man, respectively, denote galactose, glucose, *N*-acetylglucosamine, and mannose residues. Note that the N-terminal propeptide has intrachain disulfide bonds while the C-terminal propeptide has both intrachain and interchain disulfide bonds. [After Prockop, D.J., Kivirikko, K.I., Tuderman, L., and Guzman, N.A., *New Engl. J. Med.* **301,** 16 (1979).]

signal peptide is recognized by a **signal recognition particle (SRP).** The SRP binds a ribosome synthesizing a signal peptide to a receptor on the membrane [the rough endoplasmic reticulum (RER) in eukaryotes and the plasma membrane in bacteria] and conducts the signal peptide and its following nascent polypeptide through it.

Proteins bearing a signal peptide are known as **preproteins** or, if they also contain propeptides, as **preproproteins.** Once the signal peptide has passed through the membrane, it is specifically cleaved from the nascent polypeptide by a membrane-bound **signal peptidase.** Both insulin and collagen are secreted proteins and are therefore synthesized with leading signal peptides in the form of **preproinsulin** and **preprocollagen.** These and many other proteins are therefore subject to three sets of sequential proteolytic cleavages: (1) the deletion of their initiating Met residue, (2) the removal of their signal peptides, and (3) the excision of their propeptides.

Polyproteins

Some proteins are synthesized as segments of **polyproteins,** polypeptides that contain the sequences of two or more proteins. Examples include most polypeptide hormones (Section 33-3C); the proteins synthesized by many viruses, including those causing polio (Section 32-2C) and AIDS, and **ubiquitin,** a highly conserved eukaryotic protein involved in protein degradation (Section 30-6B). Specific proteases posttranslationally cleave polyproteins to their component proteins, presumably through the recognition of the cleavage site sequences. Some of these proteases are conserved over remarkable evolutionary distances. For instance, ubiquitin is synthesized as several tandem repeats **(polyubiquitin)** that *E. coli* properly cleave even though prokaryotes lack ubiquitin. Other proteases have more idiosyncratic cleavage sequences. Thus, medicinal chemists have designed and synthesized numerous inhibitors of HIV **protease** (which catalyzes an essential step in the viral life cycle) in an effort to slow the progress of, if not cure, AIDS.

B. Covalent Modification

Proteins are subject to specific chemical derivatizations, both at the functional groups of their side chains and at their terminal amino and carboxyl groups. Over 150 different types of side chain modifications, involving all side chains but those of Ala, Gly, Ile, Leu, Met, and Val, are known (Section 4-3A). These include acetylations, glycosylations, hydroxylations, methylations, nucleotidylations, phosphorylations, and ADP-ribosylations as well as numerous "miscellaneous" modifications.

Some protein modifications, such as the phosphorylation of glycogen phosphorylase (Section 17-1A) and the ADP-ribosylation of eEF-2 (Section 30-3G), modulate protein activity. Several side chain modifications covalently bond cofactors to enzymes, presumably to increase their catalytic efficiency. Examples of linked cofactors that we have encountered are N^{ε}-lipoyllysine in dihydrolipoyl transacetylase (Section 19-2A) and 8α-histidylflavin in succinate dehydrogenase (Section 19-3F). The attachment of complex carbohydrates, which occur in almost infinite variety, alter the structural properties of proteins and form recognition markers in various types of targeting and cell–cell interactions (Sections 10-3C, 11-3D, and 21-3B). Modifications that cross-link proteins, such as occur in collagen and elastin (Sections 7-2C and D), stabilize supramolecular aggregates. The functions of most side chain modifications, however, remain enigmatic.

Collagen Assembly Requires Chemical Modification

Collagen biosynthesis (Fig. 30-57) is illustrative of protein maturation through chemical modification. As the nascent procollagen polypeptides pass into the RER of the

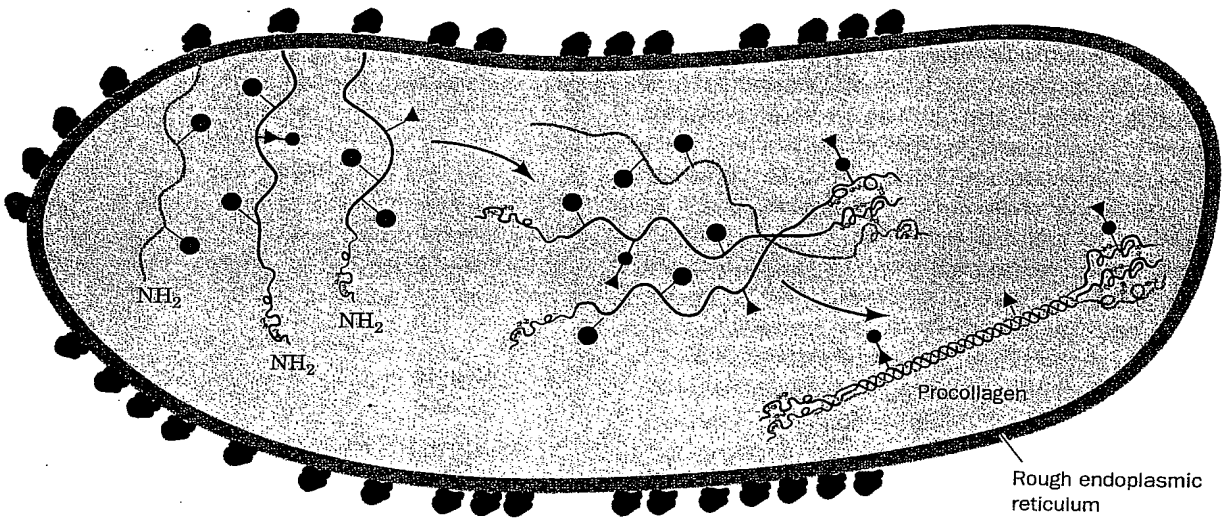


FIGURE 30-57. A schematic representation of procollagen biosynthesis. The diagram does not indicate the removal of signal peptides. [After Prockop, D.J., Kivirikko, K.I., Tuderman, L., and Guzman, N.A., *New Engl. J. Med.* **301,** 18 (1979).]